DOCUMENT RESUME

ED 157 950 95 TM 007 950

AUTHOR Echternacht, Gary; Plas, Jeanne M.
TITLE Grade Equivalent Scores. If Not Grade Equivalent Scores--Then What? NCME Measurement in Education. A Series of Special Reports of the National Council on Measurement in Education.
INSTITUTION National Council on Marine Resources and Engineering Development, Washington, D.C.
SPONS AGENCY ERIC Clearinghouse on Tests, Measurement, and Evaluation, Princeton, N.J.; National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Mar 77
NOTE 9p.
AVAILABLE FROM NCME Special Publications, 202 S. Kedzie Hall, Michigan State University, East Lansing, Michigan 48824 ($1.50 ea., 25 or more copies, $0.75 ea.)
JOURNAL CIT NCME; v8 n2 Spr 1977

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Achievement Rating; *Achievement Tests; Elementary Education; *Grade Equivalent Scores; *Norm Referenced Tests; *Scores; Scoring; Scoring Formulas; Standardized Tests; Statistical Analysis; Testing Problems; *Test Interpretation; Test Results

ABSTRACT

While most school districts believe they understand grade equivalent scores, teachers, parents, and measurement specialists frequently misinterpret this apparently simple statistical expression. Echternacht's article describes the construction, application, and interpretation of grade equivalent scores from the test publisher's perspective. Although publishers continue to provide this statistic, some urge the use of percentile standards or standard scores when evaluating and reporting test results to the public. Plas, reasoning that a test consumer should choose a statistic which represents relationships that are best understood by the user, discusses the relationships underlying the most commonly used statistics in standardized achievement testing. Of the five--raw score, T-score, stanine, percentile rank, and grade equivalent score--the latter two are the most easily misunderstood and misinterpreted. When the location of average range is mastered for each of the five statistical expressions, the test interpreter is free to decide whether an individual score represents average, below average, or above average performance. (CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED157950

Grade Equivalent Scores[1]

Gary Echternacht
Research Scientist
Educational Testing Service

If Not Grade Equivalent Scores - Then What?[2]

Jeanne M. Plas
Psychology Faculty
George Peabody College for Teachers

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Irwin J. Lehmann

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ABOUT THIS ISSUE

As ME was originally conceived, its mission was to serve the practitioner. Some might argue that we have failed to provide enough material written for such professionals. This issue represents one (of many to follow) such attempt to reach one of the ultimate consumers of our measurement wares, the classroom teacher.

Within this issue are contained two articles relating to the classroom teachers' use of test norms in their continuous attempt to describe the performances of their pupils. The grade equivalent score has been held in awe by some and with trepidation by others. It is safe to say that this score has been the most misinterpreted statistic available to us in describing pupil performance.

Dr. Gary Echternacht has written a concise article describing the construction, application, and interpretation of grade equivalent scores. Dr. Echternacht is widely published in educational and psychological measurement and is currently Research Scientist at Educational Testing Service, Princeton NJ, 08540.

Dr. Jeanne M. Plas has responded with an alternative to the use of grade equivalent scores. Dr. Plas is rapidly becoming identified as a rising star in transactional ecological psychology. She is a member of the Psychology Faculty, George Peabody College for Teachers, Nashville, TN. 37203. **PSR**

I. Grade Equivalent Scores

JAMES TOWNSHIP, N.J., Feb. 28 — Thomas Graves, Superintendent of Schools, and Samuel Hood, Principal of Horace Mann Junior High School, resigned yesterday under intense pressure from the township's school board. Board president George Rodney stated in a news conference following announcement of the resignation that the action was regrettable but that "something had to be done to shake up the mess at Horace Mann."

"I know that Dr. Graves and Mr. Hood aren't directly involved with teaching our children to read. Classroom teachers are. But administrators must take the heat for the school's poor performance. If I had my way, I would shake up the teaching staff at Horace Mann, too," Rodney said, commenting on the situation.

The controversy stems from a series of reports in the James Township *Bugle* on the school district's standardized testing program, now in its fifth year. The testing shows that students at Horace Mann are losing groung in reading. The pattern, repeated in each year of testing, reveals that over one-third of Horace Mann's students are more than one year behind in reading.

When the testing program was begun five years ago, 11 percent of the second graders at Pestalozzi Elementary School were one year behind in reading, a figure that has remained constant each year since. Although 90 percent of those students are still in the James Township school

---

[1,2]These articles were commissioned by the ERIC Clearinghouse on Tests, Measurement and Evaluation pursuant to a contract with the National Institute of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Points of view or opinions do not necessarily represent the official view or opinions of the National Institute of Education or the ERIC Clearinghouse on Test, Measurement and Evaluation.

IM007 950

system at Horace Mann, the percentage of students who are more than one year behind has increased steadily, to the point where 33 percent are now one year behind.

Grade	Percentage of Students one year behind in reading	
2	11	
3	22	
4	27	Pestalozzi
5	28	
6	29	
7	33	
8	35	Horace Mann
9	38	

Results of the school district's testing programs in 1976 show that more students are continuing to fall one year behind in reading in the higher grades. This same pattern has been found in each of the five years of testing in the school district.

What is wrong at Horace Mann? Are its teachers teaching poorly? Is the test measuring abilities that are not being taught? If we use the test scores as a guide, nothing is wrong at Horace Mann, as a matter of fact. The percentage of students who are one year behind at each grade level is exactly the same as that for the entire nation. The data do not indicate poor teaching or a lack of correspondence between the test and the curriculum. The difficulty arises because the test scores are reported in grade equivalents and are misinterpreted.

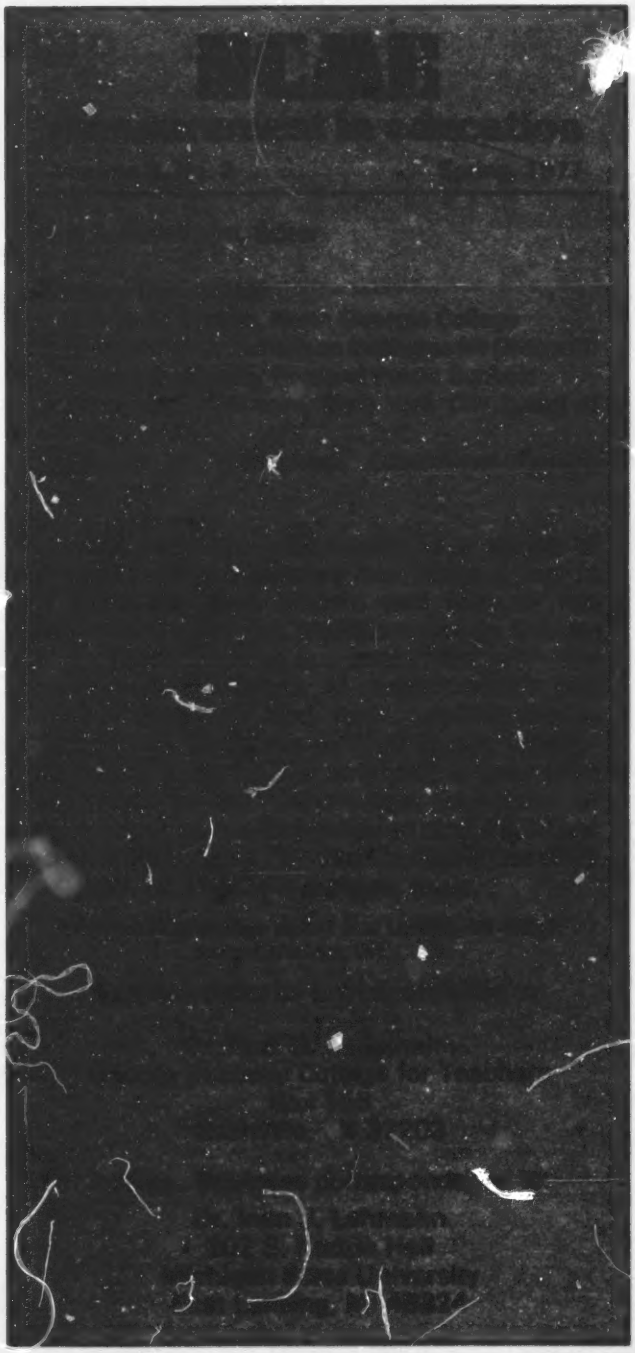
The preceding example illustrates one of the difficulties in interpreting grade-equivalent scores. They simply do not mean the same thing at different grade levels. From the example, we can see that being "one year behind" in reading at the second grade means that a child scored very low on the test — better than only 11 percent of his peers. But scoring "one year behind" at the ninth grade is not a poor performance. A student who scores one year behind in that grade performs better than 38 percent of his peers, a result that could be considered average compared with the national norm.

Another interesting aspect of grade-equivalent scores is that scoring "one year above grade level" is not interpreted in the same way as scoring "one year below grade level." For example, the above table shows that scoring one year below grade level in the ninth grade places a student at the 38th percentile — that is, the student scored better than 38 percent of his peers. If "one year above" and "one year below" are comparable, we would expect that a student in the ninth grade who scored "one year above grade level" would have 38 percent of the students in the national norm scoring higher than he or she did — that is, we would expect that score to be at the 62nd percentile. However, the test publisher's manual shows that scoring "one year above grade level" corresponds to a percentile rank of 73.

Interpreting grade equivalents

Test publishers who provide scores in grade equivalents generally warn test users of their inherent shortcomings and misinterpretations. Although publishers continue to provide grade-equivalent scores, some urge the use of such other scores as percentiles, stanines, or standard scores because, while most school districts believe they understand grade-equivalent scores, most people, including some measurement specialists, do not understand them fully.

Publishers usually will say that a grade-equivalent score should be interpreted as the score corresponding to that earned by the median, or average, student at the particular grade level. Thus, if a student scores a grade equivalent of 6.2 (sixth year, second month), we should interpret that score to mean that, on this particular test, he or she scored exactly as we would *expect* the typical student in the second month of sixth grade to score.



The word "expect" is important. The statement does not mean that these typical students *were* in the norming sample. In fact, because test scores tend to deviate from the average, the chances are that no students in the grades implied by the grade-equivalent scores ever took the test! For example, the California Achievement Test, level 3, a widely used standardized achievement test used in fourth and fifth grades, is highly regarded by measurement specialists. In the publisher's standardization program, level 3 was administered to students in grades 4, 5, and 6 during the sixth month of the school year. Nevertheless, grade-equivalent scores for reading may range from .6 to 12.9 even though no students below grade 4.6 or above grade 6.6 ever took level 3 of the California Achievement Test. No students in grades 4.7, 4.8, 4.9, or any of the grades between 5.6 and 6.6 took the test, either. Such grade equivalents are only estimates, and they may be poor estimates. If a student in fourth grade scores 3.8 on level 3 of the California Achievement Test, we can only estimate that such a score is "typical" of the average student in the eighth month of third grade; no actual student at that grade level took the test.

Perhaps the most common misinterpretation of grade-equivalent scores (especially by parents) is to consider them to be indicators of the ability to perform work required of a student at the grade level indicated by the score. A fourth-grade student who scores a 6.2 in math is believed capable of "doing sixth-grade work in math." Although in some circumstances this may be true, such a conclusion is not justified by the test scores — and of course test publishers warn against such interpretations. The fourth-grade test covers material typically taught in fourth grade. Sixth-grade students will be working on more complex areas of mathematics. A fourth-grade student may score a 6.2 grade equivalent on a standardized math test, yet be unable to do some sixth-grade work that many sixth graders who earn a score of 4.2 can do. When another fourth-grade student scores a grade equivalent of 2.4 on the same math test, it does not mean that he or she is doing "second-grade work." There are some areas of math he or she has learned while in fourth grade that the above-average student in second grade cannot handle. Conversely, there are some math areas in which our fourth grader cannot do as well as the average second grader.

Many people also become unduly concerned when a student scores "below grade level." To them it sounds much better to score one month above grade level than it does to score one month below grade level, even though there is little difference between the two scores. They regard scoring at grade level as a minimum level of achievement for normal progress rather than a simple median, where by definition one-half the norms group scored below grade level.

How publishers construct grade-equivalent scores

Many of the difficulties associated with grade-equivalent scores become evident when the publishers describe the process used to obtain them. Grade-equivalent scores are derived from the publisher's standard score scales — equal-interval scales that enable one to compare scores on a certain test at one level with scores on another test at another level. When we say that a scale is equal-interval, we mean that the differences in achievement scores mean the same everywhere on the scale. A difference of 10 points means the same at both the high and low ends of the scale. If our standard score scale ranges from 100 to 200, the growth in achievement represented by the interval from 120 to 130 indicates the same amount of growth as does the interval between 160 and 170. A ruler calibrated in inches is an equal-interval scale for measuring length. A Fahrenheit or Celsius thermometer is an equal-interval scale for measuring temperature. Publishers give their standard score scales different names — Achievement Development, Converted, Expanded, and Growth Scale Scores, for example, and each publisher's scale covers a different range of numbers.

The standard score scale functions at all levels of the publisher's test, sometimes beginning with a kindergarten test and running through a twelfth-grade test. Each grade level's score distribution is assumed to be normal or bell shaped on the scale and overlapping as indicated in Figure 1. This scale score provides a theoretical way to equate the standard scores earned on different levels of the publisher's test. Individuals with the same standard score are "equated" even though they did not take the same test. Using standard score scales, we can compare growth from year to year. If a child gained 10 scale score points from second to third grade and 15 scale score points from third to fourth grade, we would say that his growth was accelerating. We use the standard score scale to measure a child's academic growth just as we use the ruler and scale to measure a child's growth in height and weight.

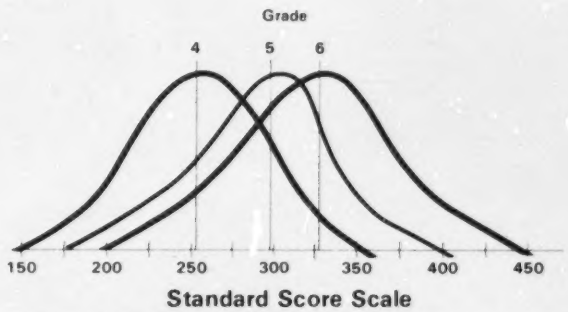


Figure 1. Score distributions for 4th, 5th, and 6th grade scores on the standard score scale.

To obtain grade equivalents from the standard scores, it is necessary to go through the following steps: a) Median standard scores are plotted for grade placement at the time or norming, as shown in Figure 2. In the figure, the fifth-grade test median standard score was 70, which was plotted for the norming grade placement (5.2 in this case). b) After all the median standard scores have been plotted, a smooth curve is drawn through the points. This curve shows the correspondence between grade-equivalent and standard scores.

Perhaps the most important aspect of Figure 2 is the fact that the line connecting the median standard scores is curved rather than straight. It begins to rise sharply, then levels off. When this happens, we lose the equal interval property of standard scores as they are converted to grade-equivalent scores. A standard score gain from 70 to 80 represents a growth of about 1.8 years (from 5.2 to 7.0), while a standard score gain from 40 to 50 represents a growth of about one year (from 1.7 to 2.7). Even though our achievement gain was the same on the standard

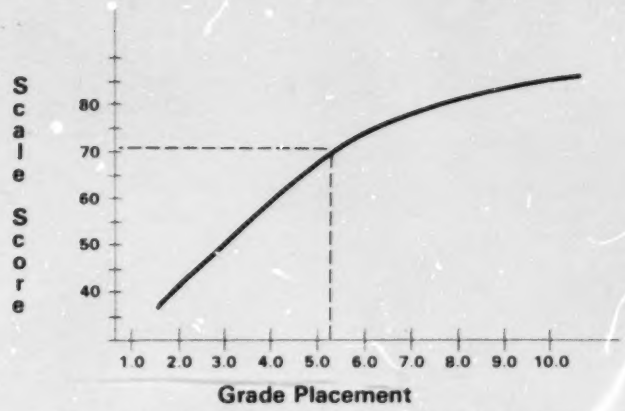


Figure 2. Conversion from standard scores to grade-equivalent scores for a hypothetical test.

score scale, it appears unequal on the grade equivalent scale.

Once the equal interval property is lost, it is no longer meaningful to perform arithmetic operations on numbers. We cannot add grade equivalent scores. We cannot calculate a meaningful class average, and program evaluations based on average class performance begin to lose their meaning.

Conclusion

Most schools receive standardized test score results in grade equivalent scores; but at least certain of the deficiencies of these grade equivalent scores for score interpretation and project evaluation are pointed out, and fortunately, test publishers provide other scores that are more useful. When inidividual scores are interpreted to students or parents, percentile scores represent a scale intuitively meaningful to most people. Percentiles, however, are not equal interval and not recommended for evaluation studies. The standard score scales provided by the publisher are equal interval and should be used when performing mathematical operations with test scores. If evaluation results are to be reported to the community and standard scores are considered too complex, one can translate the final results from standard scores to percentiles before interpreting evaluation results.

Bibliography

Horst, D.P., and Tallmadge, G.K. A procedural guide for validating achievement gains in educational projects. Monograph Series on Evaluation in Education, No. 2, Los Altos, CA., RMC Research Corp., 1976 (ED 126 135)

Horst, D.P. What's bad about grade-equivalent scores? Technical Paper No. 1, ESEA Title I Evaluation and Reporting System: Mountain View, CA., RMC Research Corp., 1976.

Hathaway, W.E. The appropriate and inappropriate uses of grade level equivalents in school evaluation. Paper presented at Annual Meeting of the American Educational Research Association, Washington, D.C., 1975 (ED 109 426)

Perry, D. Interpreting standardized test scores TR-8000. St. Paul, MN., University of Minnesota, Student Counseling Bureau, 1971 (ED 053 201)

Ricks, J.H. On telling parents about test results, Bull. 54, New York, Psychological Corporation, December 1959 (ED 079 386)

If Not Grade Equivalent Scores — Then What?

Jeanne M. Plas
Psychology Faculty
George Peabody College

It seems that measurement specialists have finally learned what interpreters of classroom tests have known for some time now — something is not quite right with grade equivalent scores. It's a problem of miscommunication and misunderstanding resulting in misuse. Several recent articles (Echternacht, 1977; Armer, 1977) have reported some of the most common misconceptions concerning the use of grade equivalent scores. It is almost disheartening to read of problem after problem associated with a score that seems, at first glance, so straightforward and usable. It is precisely the deceptively simple nature of the expressed grade equivalent score that has been responsible for its popularity as a means of reporting pupil achievement test performance. It seems so reasonable to interpret a grade equivalent score of 3.7 as meaning only that the child performed on the test in the same way as the average third grader who takes the test in March. Yet more and more consumers are beginning to know that such a statement is an over-simplified version of what really is true. Indeed, from certain points of view such a statement can be considered false.

The test consumer might well ask why the grade equivalent score was not designed to provide more directly than it does what the user is so likely to assume given the simple nature of the grade equivalent expression. In other words, why isn't it as simple as it looks? The answer to this question lies within the purpose and function of statistics in general.

Both simple and complex statistics are designed to illustrate relationships within a set of meaningful numbers. For example, comparing any given score to the arithmetic mean (a simple statistic) will indicate how the score stands in relation to a central score, the average. The purpose of any statistic ever devised can be indicated in a way similar to this example. Statistics always stand for relationships among numbers within a group or across groups. The primary function of any statistic is to distill information regarding relationships into as simple an expression as possible. Ultimately, the highest form of the art of inventing a usable statistic involves producing a simple numerical expression that can communicate a large amount of relational information. Quite appropriately, the more complex the statistic, the more relationships it represents. Thus, an apparently simple statistical expression can represent a set of complex relationships. If it represents a number of such relationships, its interpretation demands that these relationships be incorporated into its explanation.

Each statistic used in explanation of standardized achievement test results represents a different set of relationships. Thus, there is no such thing as a "best" statistic for interpretation purposes. Each statistic merely represents relationships from a different point of view. Given the wide variety of statistics available for interpretation of standardized achievement scores, a user of tests is not bound to accept any one statistic for individual interpretation purposes. The art of test score interpretation is not dependent solely upon the statistic used. It is the transaction between the statistic and the user's knowledge of how the statistic "works" that makes for

good test interpretation. *A test consumer should always interpret a score through use of that statistic which represents relationships that are best understood by the user.*

The remaining portion of this discussion is concerned with a simple description of important relationships[1] which underlie the most commonly used statistics within the field of norm-referenced[2] achievement testing. The majority of these statistical expressions are found in the technical and administrative manuals of commercially popular tests in use in schools today. Persons charged with interpreting achievement scores should choose the statistic that represents relationships which are most easily understood and meaningful to them.

The Raw Score

The raw score represents the total number of correct responses on a test or subtest. It tells very little about the person who took the test and nothing about performance in relation to others who have taken the test. Before the student's performance can be evaluated in terms of the performance of others, the raw score must be converted to one of the following expressions so that such relationships can be identified.

Percentile Rank

A table located within the test manual provides for quick conversion from raw score to percentile rank. The percentile rank indicates the percent of people scoring at or below an individual's raw score. Percentiles range from 1 to 99 with the mean percentile located at 50. Thus, a percentile rank of 65 shows that the individual's performance meets or exceeds that of 65% of the sample upon whom the test was normed. The relationships that need to be attended to when using this statistic grow out of the percentile's involvement with the normal curve ( ),[3] a bell-shaped curve which represents the frequency with which people distribute themselves across the range of possible scores. The normal curve illustrates that most people, approximately 68 percent, tend to score in the middle range while significantly fewer score at the lower and upper ends of the range. The normal curve-based percentile ranking system is designed to represent this situation and thus, the test interpreter must be aware of the following two relationships in addition to the basic definition of the percentile rank cited above.

[1]For an excellent figural description of these relationships the reader is referred to Psychological Corporation Test Service Bulletin 48 (1954) where the relationships are illustrated through visual reference to the location of standard deviations on the normal curve.

[2]Interpretation of statistics associated with criterion-referenced testing is not included in this discussion. The only distribution considered is the normal distribution since it is this distribution which is most important for interpretation of standardized educational tests.

[3]Percentiles are not dependent upon a normal distribution. However, percentiles used by teachers in connection with standardized tests have almost always been adapted to the normal curve. Thus, an understanding of norm-referenced percentiles is crucial for the classroom teacher.

[4]Throughout this discussion, the average range is identified as the range encompassed between ± one standard deviation from the mean. However, the reader should be aware that use of the standard error of measurement dictates the conceptualization of an average range that will always be smaller than that derived from the use of the standard deviation alone. While the standard deviation provides for a range about the mean, the standard error of measurement provides a range that can be used with an individual score. It is a statistic which attempts to account for error which is attributed to the measurement process itself.

The first point to be remembered is that the average range[4] on standardized achievement tests is located between the 16th and 84th percentiles. By simply subtracting 16 percent from 84 percent, the reader will discover that the 68 percent called for within the average range has been provided for through application of these two percentiles to the lower and upper limits of the average range. Thus, for example, while the 30th and 70th percentiles may appear to many to be below and above average, respectively, they are both well within the average range.

The second relationship of importance is that the distances between percentiles are not equal across the entire percentile range. A percentile of 40 is not equally distant from percentiles of 30 and 50. Thus, equal numbers of raw score points are not found from percentile to percentile. Behaviorally, this means that a child has to answer more questions correctly in order to move shorter distances within the middle range. The exact opposite is true within the above and below average ranges. In some cases, as many as five correct answers are represented between the 50th and 51st percentiles while one additional correct answer can move a child from the 90th to 95th or 1st to 5th percentile. Moving from the 50th to 55th percentile indicates less change in achievement level than does moving from the 90th to the 95th percentile.

Failure to use the above information when interpreting a percentile will always lead to erroneous conclusions. Mastery of the percentile rank concept demands that the relationships identified above be incorporated into any attempt at interpretation.

T-Score

If T-scores are suggested for use by the test publisher, a raw score to T-score conversion table is provided. Over 99 percent of all subjects will score within a T-score range of 20 through 80. The mean is located at a T-score of 50. T-scores have been designed so that the two relationships which need attention when interpreting percentiles have been accounted for statistically within the nature of this standardized scoring system. Equal distances across the T-score range can be assumed. Thus, the relationship of primary interest becomes the fact that the average group, 68 percent of the national sample, falls somewhere between T-scores of 40 and 60.

In at least one respect, T-scores provide an opportunity for clearer interpretation than do percentile ranks. The advantage is located within the initial naivete of the consumer with respect to T-scores. Since T-scores are not as popular within the common culture, there is no misinformation to be corrected as is the case with percentiles.

From another point of view, some information loss can be expected to occur through use of T-scores. For example, the fact that approximately 84 percent of those taking the test fall at or below a T-score of 60 is representative of a relationship that is not intuitively obvious. Yet such relationships need to be understood and memorized in order for the full use of the statistic to occur. The principle underlying the relationship just described is that T-scores are directly comparable to percentiles. Thus, the lower and upper limits of the average range are equivalent for both T-scores and percentiles. Since the average T-score range is 40-60, a

T-score of 40 is equivalent to a percentile rank of 16 while a T-score of 60 is equivalent to a percentile of 84. The relationship holds true across the full range of T-scores and percentiles.

Stanines

Stanines can also become readily available for interpretation through use of a raw score to Stanine conversion table located in most achievement test manuals. The word stanine is a shortened form of the term, "standard nines", which indicates that the entire normal (bell-shaped) distribution is divided into nine parts. The fifth stanine represents that group of testees which falls to the immediate left and right of the mean, and contains 20 percent of the entire sample. Although this is the middle-most group, the 5th stanine does not represent the entire average range. The average range extends from a point in the third stanine to a point in the seventh stanine. Again, approximately 68 percent of the entire sample falls within this range. The major problem when using stanines occurs when an attempt is made to interpret scores which fall in the third and seventh stanines. The lower and upper limits of the average range are imbedded within these two stanines, respectively. This means that approximately two-thirds of those who score at the seventh stanine are within the average range while approximately one-third score above the average range (and conversely for the third stanine). Thus, when only stanine information is available and the individual score falls within the third or seventh stanines, it is impossible to accurately determine performance with reference to the average range. In such cases, it is probably best to refer to such performances as borderline below average (third stanine) or borderline above average (seventh stanine).

Grade Equivalent Scores

Grade equivalent scores are the result of the conversion of raw scores into expressions which indicate achievement level in terms of academic development. Grade equivalent expressions such as 2.7 and 5.3 are to be read as second grade, seventh month and fifth grade, third month, respectively. Thus, grade equivalent scores, unlike other statistical expressions, are designed to express achievement in terms of the American public school grade level system. While statistics such as percentiles and T-scores may be rather directly applicable for use in European or Asian countries, grade equivalent scores, for the most part, are not. They are unique to the American educational system and their interpretation depends upon familiarity with that system.

The basic relationship interpretable through use of a grade equivalent score is the relationship between the individual's actual grade placement and the mean grade equivalent score obtained by the group which shares that actual grade placement. For example, in constructing grade equivalent scores it is predetermined that the mean score obtained by the sample group of second grade, seventh month testees will be identified as a grade equivalent score of 2.7. Put another way, if during the norming process the test is administered to second grade children during March (the seventh month), then the average raw score obtained by these children will convert to a 2.7 grade equivalent score. The major problem in interpreting grade equivalent scores occurs when one overlooks the fact that this mean grade equivalent score only represents the average or center score of the entire distribution. All children are not expected to score at the mean. Approximately 68 percent are expected to score around this mean. Thus, individual scores are only interpretable as they relate to a range of grade equivalent scores that represents the average group. An appreciation for the severity of the interpretation problem increases when one becomes familiar with the fact that grade equivalent average ranges or bands are very different across subtests and across grade levels.

The manual of one widely used achievement test reports that the following ranges indicate average peformance (± one standard deviation) across the identified subtests at the third grade level (end of Year Testing.):

Subtest	Average Grade Equivalent Range
Reading	2.3 to 5.3
Language	2.5 to 5.7
Mathematics Computation	3.0 to 4.8

The grade equivalent scores of 5.3, 5.7, and 4.8 all fall at the 84th percentile just as those identified as lower limits (2.3, 2.5, 3.0) fall at the 16th percentile. The average range varies across subtests and is more extended than might be expected. Grade placement scores of 5.0, 5.1, and 4.3 across the subtests might seem, at first glance, to indicate superior performance for a third grade child. Yet they are well within the average range. This example is not atypical. Rather, it represents the prevailing situation.

It is important to note also that, in the above example, grade equivalent scores of 5.3, 5.7, and 4.8 on reading, language and computation subtests respectively, are all representative of performances which equal or exceed that of 84 percent of the third grade children on whom the test was standardized. The performances are thus equivalent in interpretation even though represented by different grade placement scores. The easiest and best interpretation of grade equivalent scores probably is made through use of the percentile rank at which they are located.

The problem which occurs when interpreting across grade level also deserves special attention. It is located in the fact that at the lower grade levels the average range is more compressed than it is at the upper grade levels. A child who scores one grade level below the mean in the first grade in any skill area may well have scored below the average range. This is not true at the upper grade levels where such a performance (one year below grade level) may well be solidly within the average range. At the upper grade levels, a larger amount of material is covered: More years of achievement are being assessed.

Most problems associated with grade equivalent interpretation are located in a misunderstanding of the variable nature of the grade equivalent average range across subtests and across actual grade placements. Once the necessity for consulting such ranges is realized, the test consumer often will be able to locate these ranges within test manuals. This information is becoming increasingly more available as test producers become

acquainted with the problems the test consumer experiences in trying to use only common sense when interpreting grade equivalent scores. As is the case with percentile ranks, common sense often is not enough when correct interpretation is the goal.

Another area of misconception specifically involves the interpretation of above average grade equivalent scores. A grade equivalent score of 7.2 on a fourth grade arithmetic subtest does *not* mean that the child is capable of performing seventh grade mathematics work. Grade equivalent scores are derived from the relationship of the student's performance to norm-referenced percentiles. They are not derived by incorporating into the test significantly above grade level work. Thus, in the example above, the fourth grade student was not asked to solve seventh grade math problems. Without giving a seventh grade test, we have no valid way of knowing how the student performs in relation to the seventh grade work.

The statistics cited above are those most commonly used in connection with standardized achievement tests. The statistic of choice always ought to be that statistic which represents relationships that are best understood by the test consumer. It is unfortunate and almost paradoxical that the two statistical expressions which have most relevance to the common culture, the percentile rank and grade equivalent score, are the two which are most easily misunderstood and misused. It is perhaps their apparent culture-relevance which is at the root of the problem. Because test consumers are familiar with the terms, they are quick to conclude that they understand the basic relationships represented by the terms. Since these relationships are intimately involved with the conception of a normal distribution, it is not always true that the relationships can be explained through use of familiar concepts and common sense. Statistical relationships must be learned and memorized if errors are to be avoided.

The person who experiences difficulty in learning and using the concepts identified above often feels incompetent due to a lack of knowledge concerning where to begin given what appears to be a massive amount of new and unfamiliar information. Such a test consumer is urged to consider a beginning that involves the concept of the average range. Each statistic associated with achievement testing is best used if the average range is located and individual scores are then related to it. Although average ranges look quite different from statistic to statistic, equivalent information is supplied. All average ranges discussed here encompass the middle 68 percent of test performances. The following table identifies the average range for those statistics referred to in this discussion:

Statistic	Average Range
Percentile	16% tile — 84% tile
T-score	40 — 60
Stanine	mid 3rd — mid 7th
Grade Equivalent Score	Highly variable across tests, tests, subtests, and grade levels. Test Manual must be consulted.

Once the location of the average range is mastered for each statistical expression, the test interpreter is free to decide if an individual score represents average, below average, or above average performance. Further interpre-

REPORTS AVAILABLE

Back issues at *Measurement in Education* are available for 1.50 each per single issue, and for 75¢ each in quantities of 25 or more for a single issue.

Vol. 1, No. 1	*Helping Teachers Use Tests* by Robert L. Thorndike
No. 2	*Interpreting Achievement Profiles — Uses and Warnings* by Eric F. Gardner
No. 3	*Mastery Learning and Mastery Testing* by Samuel T. Mayo
No. 4	*On Reporting Test Results to Community Groups* by Alden W. Badal & Edwin P. Larsen
Vol. 2, No. 1	*National Assessment Says* by Frank B. Womer
No. 2	*The PLAN System for Individualizing Education* by John C. Flanagan
No. 3	*Measurement Aspects of Performance Contracting* by Richard E. Schutz
No. 4	*The History of Grading Practices* by Louise W[illegible]er Cureton
Vol. 3, No. 1	*Using Your Achievement Test Score Reports* by Edwin Gary Joselyn & Jack C. Merwin
No. 2	*An Item Analysis Service for Teachers* by Willard G. Warrington
No. 3	*On the Reliability of Ratings of Essay Examinations* by William E. Coffman
No. 4	*Criterion-Referenced Testing in the Classroom* by Peter W. Airasian and George F. Madaus
Vol. 4, No. 1	*Goals and Objectives in Planning and Evaluation: A Second Generation* by Victor W. Doherty and Walter E. Hathaway
No. 2	*Career Maturity* by John O. Crites
No. 3	*Assessing Educational Achievement in the Affective Domain* by Ralph W. Tyler
No. 4	*The National Test-Equating Study in Reading (The Anchor Test Study)* by Richard M. Jaeger
Vol. 5, No. 1	*The Tangled Web* by Fred F. Harcleroad
No. 2	*A Moratorium? What Kind?* by William E. Coffman
No. 3	*Evaluators, Educators, and the Publics: A Detente?* by William A. Mehrens
No. 4	*Shall We Get Rid of Grades?* by Robert L. Ebel
Vol. 6, No. 1	*On Evaluating a Project: Some Practical Suggestions* by John W. Wick
No. 2	*Measuring Reading Achievement: A Case for Criterion — Referenced Testing and Accountability* by S. Jay Samuels and Glenace E. Edwall
No. 3	*Security in a Citywide Testing Program* by Anthony J. Polemeni
No. 4	*Dear Mama: Why don't they love me anymore?* by Thomas J. Fitzgibbon
Vol. 7, No. 1	*The Discrepancy Evaluation Model (Part I)* by Andres Steinmetz
No. 2	*The Discrepancy Evaluation Model (Part II)* by Andres Steinmetz
No. 3	*What Graduate and Professional School Students Think About Admission Tests* by Leonard L. Baird
No. 4	*The Paradox of Education Testing* by Robert L. Ebel
Vol. 8, No. 1	*FACE UP: A Framework for Assessing Career Educators Utilization Programs* by Bruce W. Tuckman

tation of the score depends upon mastery of the unique relationships represented by the various statistical expressions. Some of the most basic of these have been described above. Each test interpreter is free to choose a statistical expression that seems most understandable and meaningful. It is through use of such a process that consumers answer for themselves questions such as, "If not grade equivalent scores — then what?"

REFERENCES

Armer, William. *Thoughts on Achievement Testing:* A checklist for teachers. (Unpublished manuscripts ESEA Title III, Wilson County Tennessee, 1977).

Echternacht, Gary. Educational Testing Service. *Grade Equivalent Scores.*

Psychological Corporation Test Service Bulletin, *48. Methods of Expressing Test Scores.* New York: The Psychological Corporation, 1954.

SPECIAL ORDER FORM

Name: _____

Mailing Address: _____

_____

City State ZIP

_____ **JEM, V.13, #1: TEST BIAS ISSUE**

	No.	Cost
SINGLE COPY @ $4.00*	_____	_____
2-49 copies @ $3.75/copy	_____	_____
50-99 copies @ $3.50/copy	_____	_____
100+ copies @ $3.25/copy	_____	_____

_____ **V.14, #2: LATENT TRAIT MODELS ISSUE**

	No.	Cost
SINGLE COPY @ $5.00*	_____	_____
2-49 copies @ $4.75/copy	_____	_____
50-99 copies @ $4.50/copy	_____	_____
100+ copies @ $4.00/copy	_____	_____

*NOTE: ALL PRICES ARE FOR PRE-PAID ORDERS ONLY. THESE PRICES FOR MEMBERS AND SUBSCRIBERS ONLY. NON-MEMBERS ADD $2.00 TO EACH LISTED PRICE. NCME WILL ADD A $1.00 CHARGE TO ALL ORDERS REQUIRING AN INVOICE.

Please return your order form to: NCME Special Publications
202 S. Kedzie Hall
Michigan State University
East Lansing, Mi. 48824

NCME NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION

202 S. Kedzie Hall
Michigan State University
East Lansing, Michigan 48824

Second class postage paid at East Lansing, Mich. 48823 and additional mailing points.